

## Model Archive Summary for Suspended-Sediment Concentration at U.S. Geological Survey Station 11337190; San Joaquin River at Jersey Point, California

This model archive summary describes the suspended-sediment concentration (SSC) model developed to compute a 15-minute SSC time-series for the period of record: October 1, 2011 to October 2, 2014. This is the first suspended-sediment model developed for the site. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in the Office of Surface Water/Office of Water Quality Technical Memorandum and USGS Techniques and Methods, book 3 chapter 4 (USGS, 2016; Rasmussen and others, 2009). This summary and model archive are in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (USGS, 2014).

### Site and Model Information

Site number: 11337190

Site name: San Joaquin River at Jersey Point, California (SJJ)

Location: Lat 38°03'08", long 121°41'16" referenced to North American Datum of 1927, Contra Costa County, CA, Hydrologic Unit 18040003

Equipment: A YSI 6-series sonde began logging turbidity with a model 6136 sensor on December 1, 2009 and was removed on October 2, 2014.

Model number: 11337190.SSC.WY12.1

Model calibration data period: January 27, 2012 – September 12, 2014

Model application date: October 1, 2011 – October 2, 2014

Computed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)

Reviewed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)

### Physical Sampling Details and Sediment Data

All sediment data were collected using USGS protocols and are stored in the National Water Information System (NWIS) database: <https://waterdata.usgs.gov/nwis> (USGS, 2006). Discrete, boat-based samples were collected seasonally, spanning the range of site conditions and specifically targeting large sediment transport events.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999) and USGS (2006). Samples were collected using the Equal Discharge Increment (EDI) method, as velocities were not always isokinetic due to the tidal nature of the site (from Table 4-5 of TWRI09A4; USGS, 2006). A boat-based discharge measurement was collected with an Acoustic Doppler Current Profiler (ADCP) immediately before EDI sampling to determine the location of each of the five verticals. A Federal Interagency Sedimentation Project (FISP) US D-96 bag sampler was used to collect depth-integrated samples at each vertical. The channel cross section can approach 53 feet deep in the thalweg with a mean depth of approximately 43 feet. Velocities during the model calibration data period ranged from -2.94 ft/s to +3.31 ft/s. Sediment at this station is mostly fines (94% fines on average) and any potential sampling bias due to non-isokinetic sampling is considered minimal.

Samples were analyzed for SSC (mg/L) by the filtration method at the USGS Sediment Laboratory in Santa Cruz, California. All samples were also analyzed for the percentage of fines (<0.063 mm), which

can be used to identify outliers. Two EDI samples, collected on January 27, 2012 and February 22, 2012, were composited. Each of the five verticals from the remaining EDI samples were analyzed individually by the lab for quality control purposes. The average SSC from these five verticals was computed and used in the calibration dataset. Sediment results are publicly available on NWIS.

All sediment data were reviewed and approved in the USGS NWIS Water-Quality System database (QWDATA) before being applied in the calibration model.

## Surrogate Data

Continuous, 15-minute turbidity data, reported in Formazin Nephelometric Turbidity Units (FNU) and hourly, tidally-filtered discharge data (QFT), reported in cubic feet per second (cfs), were evaluated as explanatory variables for SSC. Turbidity and QFT time-series data were collected by the USGS California Water Science Center and are located at: [https://waterdata.usgs.gov/ca/nwis/uv/?site\\_no=11337190](https://waterdata.usgs.gov/ca/nwis/uv/?site_no=11337190). Turbidity data were analyzed and approved per USGS guidelines (Wagner and others, 2006). QFT data were computed, reviewed and approved before using in the sediment calibration model. Methods to compute discharge (and thus tidally-filtered discharge) follow Levesque and Oberg (2012).

## Model Calibration Dataset

The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair surrogate data with discrete sediment data (Domanski and others, 2015). Turbidity and QFT values were paired with each sediment sample with a matching window of  $\pm 15$  minutes and  $\pm 30$  minutes, respectively. The SAID manual is available at: <https://pubs.er.usgs.gov/publication/ofr20151177>.

The sample collected on June 3, 2014 did not have a corresponding turbidity value due to deletions in the turbidity time-series and could therefore not be included in the calibration dataset.

The sample collected on June 26, 2013 had one vertical rejected as being erroneous. The sample average was manually calculated from the remaining four verticals to be used in the calibration dataset.

The final calibration dataset is compiled from 15 concurrent measurements of SSC and turbidity. The inclusion of hourly QFT data in the multiple linear regression model reduced the number of observations to 14 due to a data gap in the time series. Negative QFT values cannot be log transformed, which reduced the number of observations in the multi-log model to 10. Summary statistics and the complete model calibration dataset are provided in the following sections.

## Model Development

Simple linear regression (SLR) models and multiple linear regression (MLR) models were assessed using methods described in Helsel and Hirsch (2002). Four models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2)  $\log_{10}$ -transformed model with one explanatory variable (turbidity), Model 3) linear model with two explanatory variables (turbidity and QFT) and Model 4)  $\log_{10}$ -transformed model with two explanatory variables (turbidity and QFT).

Diagnostic statistics and plots for model review were output using a combination of Matlab, SAID and the R environment (R Core Team, 2018). Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate regression models. The best model was chosen based on residual plots, coefficient of determination ( $R^2$ ), root-mean-squared error (RMSE), mean square prediction error

(MSPE), significance tests (p-values) and prediction error sum of squares (PRESS) statistics. RMSE and PRESS statistics cannot be used to compare regressions with different response variable units, so  $R^2$ , MSPE values and residual plots were used as the main determinants of model strength when comparing  $\log_{10}$ -transformed and untransformed models. Values for these statistics were computed for four models and are included in the table below. The best SLR model is a linear model with turbidity as the surrogate (highlighted in table below).

QFT was not considered further as an explanatory variable because: 1) QFT was not significant in either MLR model (p-value > 0.05), 2) the MLR models contain 14 or 10 observations, though a total of 48 samples is recommended when a second explanatory variable is included (USGS, 2016) and 3) including QFT in the final model would limit the computed time-series to an hourly record rather than a 15-minute record.

No.	$R^2$	$R^2_a$	RMSE	PRESS	MSPE	n	Type
Model 1	0.79	0.77	2.97	147	17.1	15	linear
Model 2	0.70	0.67	0.08	0.12	19.1	15	log
Model 3	0.83	0.80	2.75	160	15.4	14	multi-linear
Model 4	0.77	0.71	0.07	0.05	16.8	10	multi-log

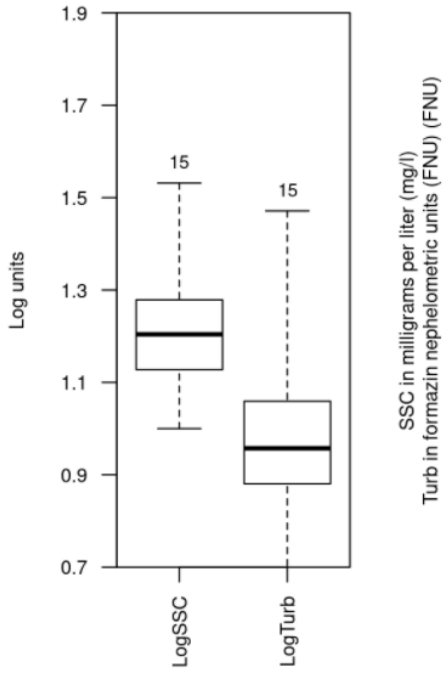
Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals from the models were inspected for values greater than three or less than negative three; values outside this range are considered potential extreme outliers. The studentized residuals were reviewed from the output reports and none of the samples were deemed to be extreme outliers. All 15 observations were retained in the model.

## Plots

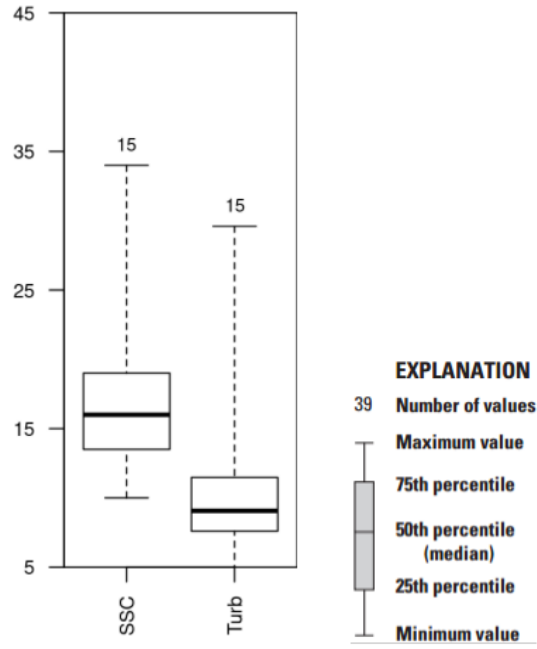
The following plots were generated using a R-based application (Version 1.0) developed by Patrick Eslick of the USGS Kansas Water Science Center. It is available at:

<http://ksWSC.cr.usgs.gov:3838/peslick/ModelArchiveSummary/>.

Boxplots of turbidity, QFT and SSC data show the range of measured data for each parameter. The third set of boxplots show SSC residuals of the SLR model by month and water year.

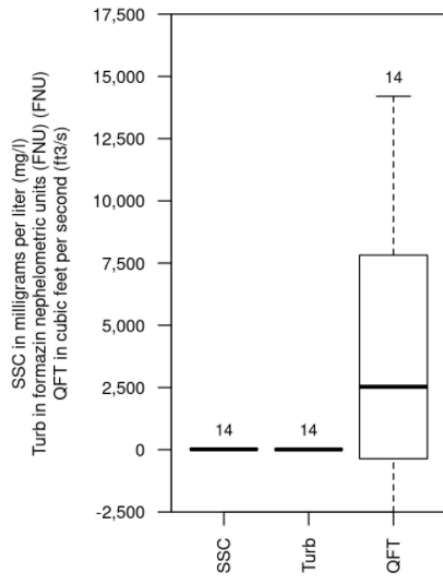


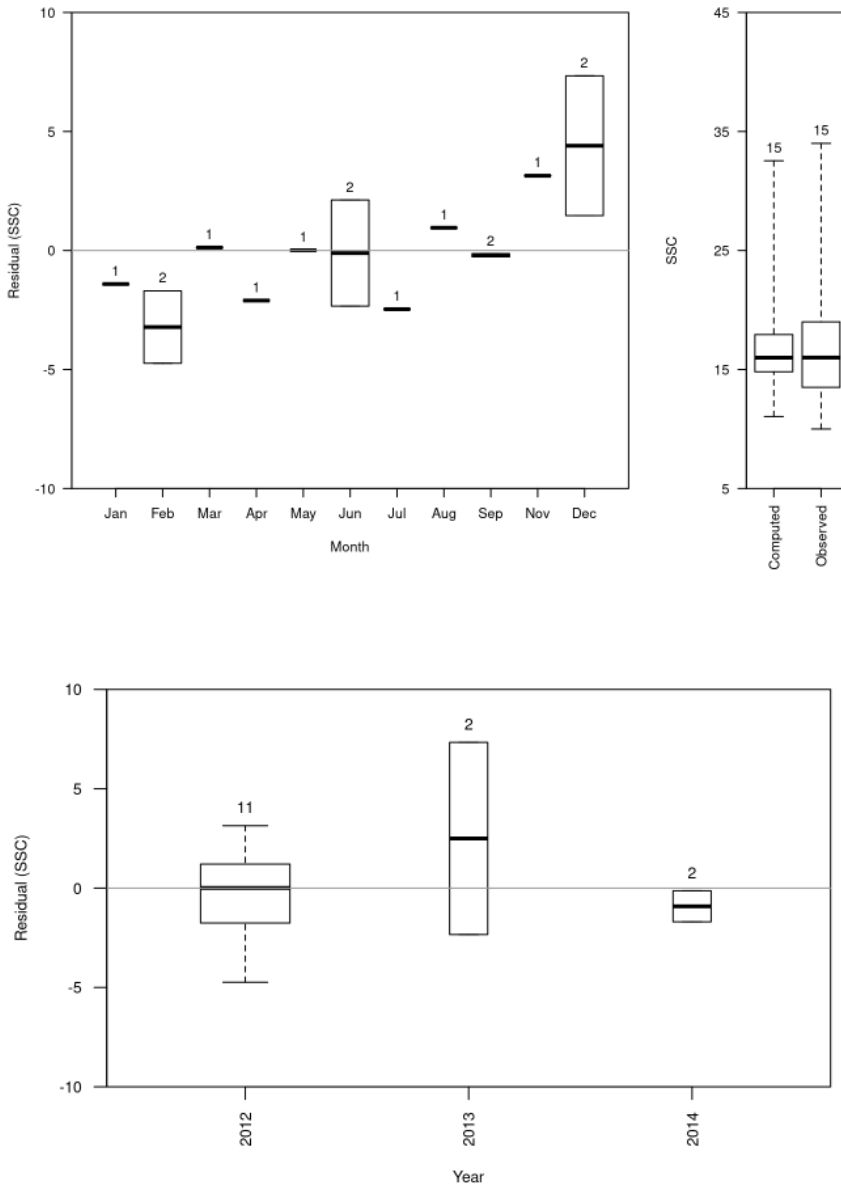
SSC in milligrams per liter (mg/l)  
Turb in formazin nephelometric units (FNU) (FNU)



#### EXPLANATION

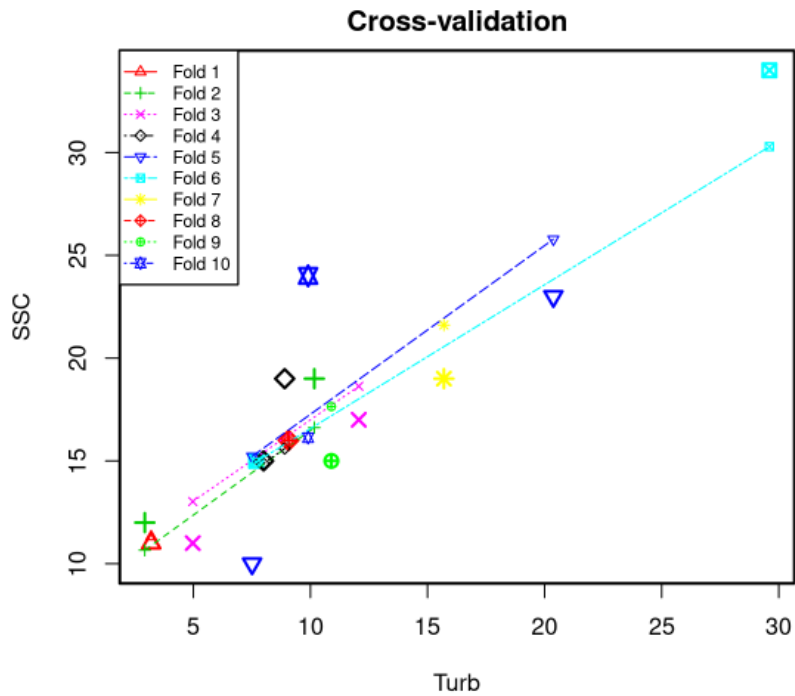
- 39 Number of values
- Maximum value
- 75th percentile
- 50th percentile (median)
- 25th percentile
- Minimum value



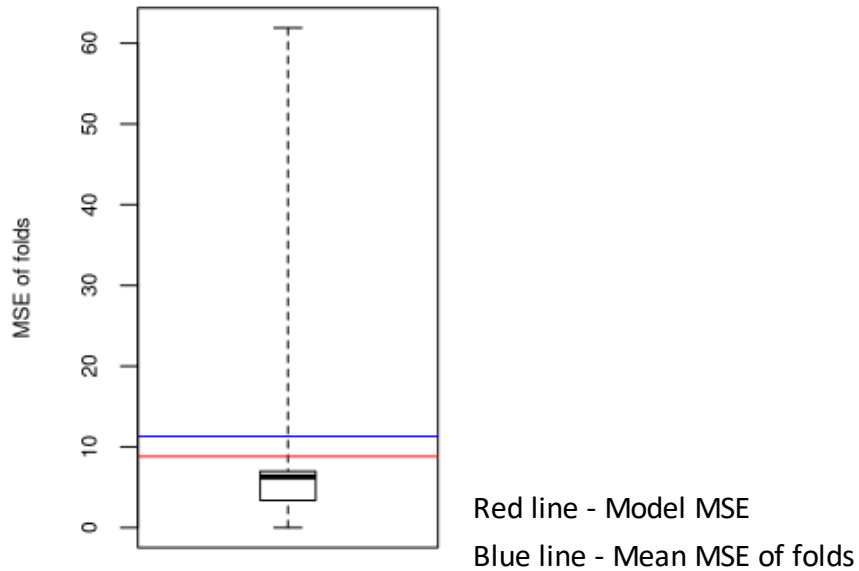


## Cross Validation

The cross-validation plot below shows a k-fold validation with k=10 for the final model. The points represent observations that were left out of each fold.



Minimum MSE of folds: 7.04E-05  
 Mean MSE of folds: 11.3  
 Median MSE of folds: 6.28  
 Maximum MSE of folds: 61.9  
 (Mean MSE of folds) / (Model MSE): 1.28



## Model Summary

The final SSC model at SJJ is a linear SLR model based on 15 concurrent measurements of SSC and turbidity collected over approximately three water years. The model is shown below with basic model information, regression coefficients, correlation and summary statistics.

Linear Regression Model	Coefficient of Determination ( $R^2$ )
$SSC = 8.69 + 0.806 * Turb$	0.789

where

*SSC* = suspended-sediment concentration, in milligrams per liter (mg/L) and

*Turb* = turbidity, in formazin nephelometric units

The SSC time-series is computed from USGS turbidity data. Minimum and maximum turbidity values for the model application period are listed below. SSC time-series data exceeding extrapolation limits were removed. This model cannot be used to extrapolate more than 10% above or below the range of samples in the calibration dataset (USGS, 2016). The extrapolated, maximum computed SSC for this model is 37 mg/L. The original maximum, computed SSC was 145 mg/L.

Parameter	Minimum	Maximum
Computed SSC (mg/L)	9.25	37
Turbidity (FNU)	0.7	170

### Suspended-Sediment Concentration Record

The SSC record is computed using this regression model on the USGS National Real-Time Water Quality (NRTWQ) website. The complete record can be found at: <https://nrtwq.usgs.gov/ca>.

**Model**

SSC = 8.69 + 0.806Turb

**Variable Summary Statistics**

	Turb	SSC
Minimum	2.93	10
1st Quartile	8	12.75
Median	9.07	16
Mean	10.73	17.33
3rd Quartile	11.77	19
Maximum	29.60	34

**Basic Model Statistics**

Number of observations	15
Root Mean Squared Error (RMSE)	2.97
Model Standard Percentage Error (MSPE)	17.1
Coefficient of determination ( $R^2$ )	0.789
Adjusted $R^2$	0.773

**Explanatory Variables**

	Coefficients	Standard Error	t value	Pr(> t )
(Intercept)	8.689	1.46	5.96	4.74E-05
log10Turb	0.806	0.12	6.97	9.73E-06

**Correlation Matrix**

	Intercept	E.vars
Intercept	1.000	-0.85
E.vars	-0.85	1.000

**Outlier Test Criteria**

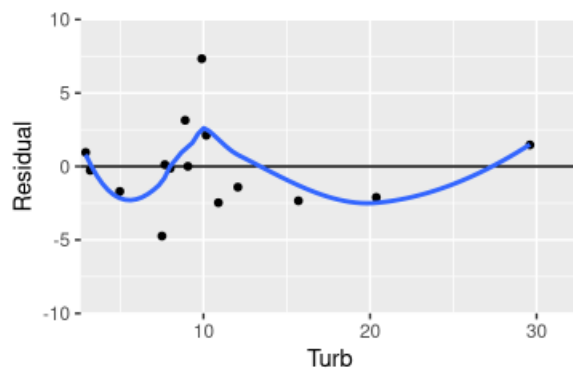
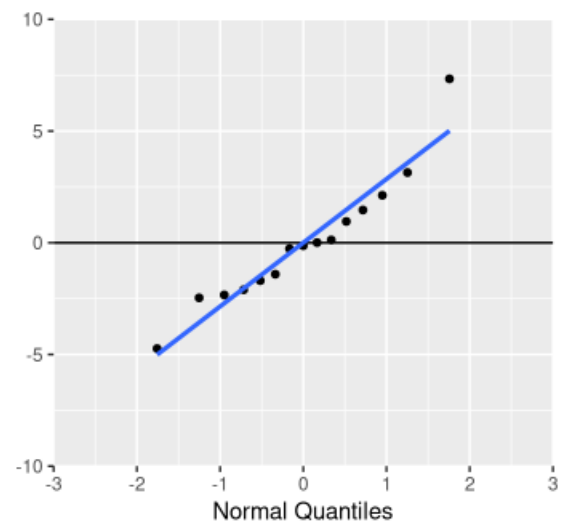
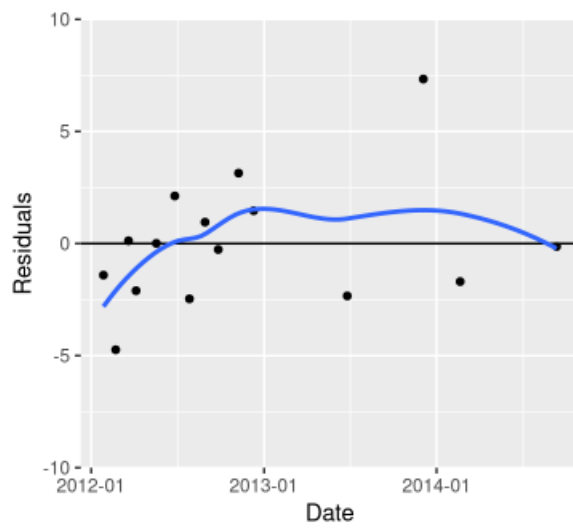
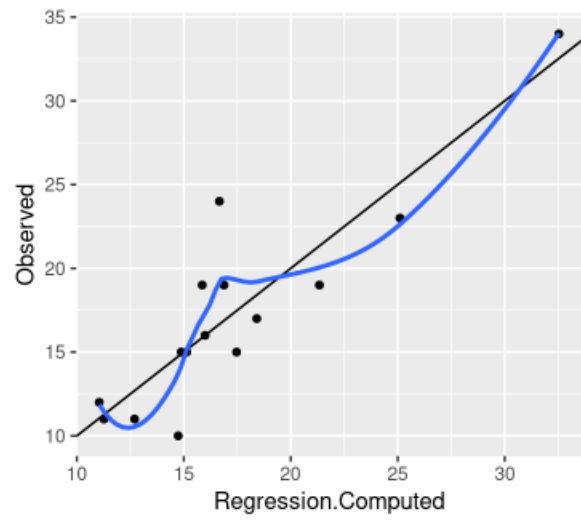
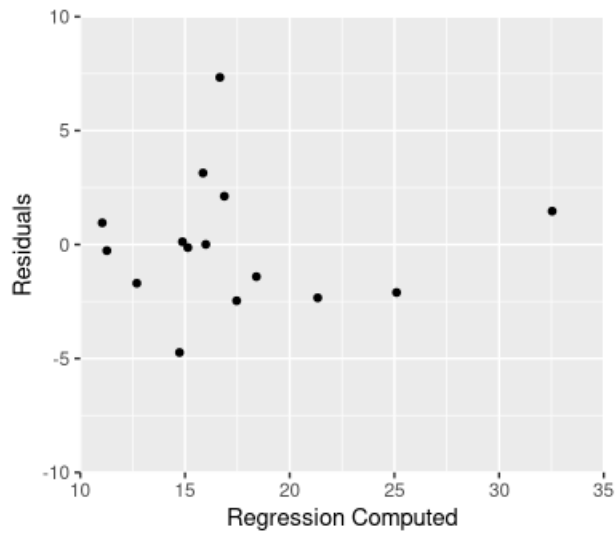
Leverage	Cook's D	DFFITS
0.4	0.192	0.730

**Flagged Observations**

Date	Time	SSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
12/10/2012	12:25	34	32.5	1.47	0.789	0.774	0.606	0.474	0.959
12/4/2013	13:40	24	16.7	7.34	2.56	3.49	0.0677	0.238	0.94

**Residual diagnostic plots**

Plots were generated using the model archive summary application developed by Patrick Eslick of the USGS Kansas Water Science Center.



Model-Calibration Dataset

	Date & Time	SSC	Turb	Computed SSC	Residual	Normal Quantiles	Censored Values
0							
1	1/27/2012 10:26	17	12.1	18.4	-1.41	-0.336	--
2	2/22/2012 12:16	10	7.51	14.7	-4.74	-1.76	--
3	3/20/2012 12:09	15	7.68	14.9	0.123	0.336	--
4	4/5/2012 11:18	23	20.4	25.1	-2.1	-0.716	--
5	5/18/2012 10:22	16	9.07	16	0.0078	0.166	--
6	6/26/2012 12:09	19	10.2	16.9	2.12	0.95	--
7	7/27/2012 10:28	15	10.9	17.5	-2.47	-1.25	--
8	8/29/2012 11:39	12	2.93	11	0.954	0.517	--
9	9/26/2012 10:24	11	3.2	11.3	-0.267	-0.166	--
10	11/8/2012 13:18	19	8.9	15.9	3.14	1.25	--
11	12/10/2012 12:25	34	29.6	32.5	1.47	0.716	--
12	6/26/2013 8:28	19	15.7	21.3	-2.34	-0.95	--
13	12/4/2013 13:40	24	9.9	16.7	7.34	1.76	--
14	2/20/2014 13:11	11	4.98	12.7	-1.7	-0.517	--
15	9/12/2014 11:21	15	8	15.1	-0.134	0	--

## Definitions

SSC: Suspended sediment concentration (SSC) in mg/l (80154)

Turb: Turbidity in FNU (63680)

## References

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., <https://pubs.usgs.gov/of/2015/1177/ofr20151177.pdf>.

- Edwards, T.K. and Glysson, G.D., 1999, Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. C2, 89 p., Available from: [https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI\\_3-C2.pdf](https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf).
- Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.
- Levesque, V.A., and Oberg, K.A., 2012, Computing discharge using the index velocity method: U.S. Geological Survey Techniques and Methods 3-A23, 148 p. Available from: <https://pubs.usgs.gov/tm/3a23/>.
- R Core Team, 2018, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Available from: <https://www.R-project.org/>.
- Rasmussen, P.P., Gray J.R., Glysson G.D., Ziegler A.C., 2009, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods book 3, chap. C4, 53 p., Available from: <https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf>.
- [USGS] U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations. book 9, chap. A4. Available from: [https://pubs.usgs.gov/twri/twri9a4/twri9a4\\_Chap4\\_v2.pdf](https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf).
- [USGS] U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: <https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf>.
- [USGS] U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, Available from: <https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf>.
- Wagner RJ, Boulger RW, Oblinger CJ, Smith BA, 2006, Guidelines and standard procedures for continuous water quality monitors: station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3, Available from: <https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf>.